PB2002-108726

COMMONWEALTH OF PENNSYLVANIA
DEPARTMENT OF TRANSPORTATION

PENNDOT RESEARCH

PENNDOT

TRAFFIC VOLUME MONITORING RELATED RESEARCH II

University-Based Research, Education, and Technology Transfer Program

AGREEMENT NO. 359704, WORK ORDER 108

FINAL REPORT

June 2002

By L. J. French, W. Iskander, and M. Jaraiedi

PENNSTATE

1855

Pennsylvania Transportation Institute

The Pennsylvania State University
Transportation Research Building
University Park, PA 16802-4710
(814) 865-1891 www.pti.psu.edu

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-PA-2002-019-97-04 (108)		

4. Title and Subtitle	5. Report Date / 6. Performing Organization Code
Traffic Volume Monitoring Related Research II	5. Report Date: June 10, 2002
	6. Performing Organization Code
7. Author(s) L. James French, Wafik Iskander, and Majid Jaraiedi	**8. Performing Organization Report No.** PTI 2002-36
9. Performing Organization Name and Address The Pennsylvania Transportation Institute Transportation Research Building The Pennsylvania State University University Park, PA 16802-4710	**10. Work Unit No. (TRAIS)**
	11. Contract or Grant No. 359704, Work Order 108
12. Sponsoring Agency Name and Address The Pennsylvania Department of Transportation Bureau of Planning and Research Commonwealth Keystone Building 400 North Street, 6th Floor Harrisburg, PA 17120-0064	**13. Type of Report and Period Covered** Final Report, 5/2/2001 - 7/1/2002
	14. Sponsoring Agency Code

15. Supplementary Notes
COTR: Chris Allison
Phone: (717) 783-9972

16. Abstract

Three objectives were undertaken as part of this research. First, a literature review and national survey of growth factor estimation procedures was conducted. Of particular interest were the frequency, duration, and spatial density of counts collected to support the estimation of growth factors. It was determined that PENNDOT might improve their growth factor estimation program through the collection of traffic data at more locations for a shorter (than three weeks) duration. Second, historical PENNDOT traffic data were used to test the impact of reducing the duration of control counts from three weeks and to determine the benefits of including more locations. It was found that the three-week duration provided little if any benefit over 48-hour counts. In contrast, the additional locations improved the growth factor estimates. Finally, historical PENNDOT traffic data were used to cluster the existing permanent continuously operating automated traffic recorders (ATRs) according to monthly factors and other criteria. It was determined that the existing Traffic Pattern Groups developed by PENNDOT should be retained.

17. Key Words	18. Distribution Statement
Traffic Data, Statewide Traffic Data Collection, Traffic Volume Monitoring, Growth Factors, Automated Traffic Recorders, Cluster Analysis, Traffic Seasonality	No restrictions. This document is available from the National Technical Information Service, Springfield, VA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified		

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

TRAFFIC VOLUME MONITORING RELATED RESEARCH II

University-Based Research, Education, and Technology Transfer
Agreement No. 359704
Work Order 108

FINAL REPORT

Prepared for

Commonwealth of Pennsylvania
Department of Transportation

By

L. James French, Ph.D., P.E.
Wafik Iskander, Ph.D., P.E.
Majid Jaraiedi, Ph.D.

West Virginia University
Morgantown, WV

The Pennsylvania Transportation Institute
The Pennsylvania State University
Transportation Research Building
University Park, PA 16802-4710

June 10, 2002

PTI 2002-36

TABLE OF CONTENTS

Page

CHAPTER 1 - INTRODUCTION 1

CHAPTER 2 - LITERATURE REVIEW AND NATIONAL SURVEY 2

2.0 Introduction 2
2.1 Literature Review 2
2.2 National Survey of Growth Factor Estimation Programs 5
2.3 Summary 12

CHAPTER 3 - TESTING OF SPATIAL VERSUS TEMPORAL COUNTING PROGRAMS . 13

3.0 Introduction 13
3.1 Methodology 13
3.2 Results 15
3.3 Conclusions and Recommendations for the Growth Factor Estimation Process 18
3.4 Limitations of the Research 19
3.5 Summary of Recommended Modifications to the Growth Factor Estimation Process . . 20

CHAPTER 4 - CLUSTER ANALYSIS OF ATRS 21

4.0 Introduction 21
4.1 Cluster Analysis 21
4.2 Conclusions and Recommendations 30

CHAPTER 5 - CONCLUSIONS AND RECOMMENDATIONS 32

5.0 Introduction 32
5.1 Literature Review and Nationwide Survey 32
5.2 Testing of Spatial versus Temporal Counting Programs 32
5.3 Cluster Analysis of ATRs 33

REFERENCES 34

APPENDIX A - Technical Material Supporting the Literature Review and Nationwide Survey A-1

APPENDIX B - Technical Material Supporting the Testing of Spatial versus Temporal Counting Programs B-1

APPENDIX C - Technical Material Supporting the Cluster Analysis of ATRs C-1

LIST OF TABLES

	Page
Table 2-1 - Continuous ATRs by State	7
Table 2-2 - Road Mileage per ATR by States in Rank Order	8
Table 2-3 - Number of Short-Term Count Sites	9
Table 2-4 - Duration of Short-Term Counts	9
Table 2-5 - Count Frequency	10
Table 2-6 - Number of Years Considered in Growth Factor Estimation	12
Table 3-1 - Allegheny County ATR Information	15
Table 3-2 - Control Count Information	16
Table 3-3 - Allegheny County HPMS Growth Rate Information	16
Table 3-4 - Absolute Error of the Control Count Growth Estimation	16
Table 3-5 - Comparison of Average Absolute Errors of the Various Methods	17
Table 3-6 - Results of Count Duration Comparison	18
Table 4-1 - Clustering Process - Unconstrained by Functional Class	23
Table 4-2 - Clusters Unconstrained by Functional Class	24
Table 4-3 - Results Based on Unconstrained Clustering and Arbitrary Adjustment	25
Table 4-4 - Values Assigned to Functional Classification Codes	26
Table 4-5 - Clustering Process with Functional Class Used as a Parameter	27
Table 4-6 - Clusters Constrained by Functional Class	27
Table 4-7 - Comparison of Clusters and Existing Traffic Pattern Groups	28
Table 4-8 - Rural Minor Arterial Clustering Process	29
Table 4-9 - Rural Collector Clustering Process	29

Table 4-10 - Combined Rural Minor Arterial and Collector Clustering Process 30

Table 4-11 - Precision Analysis for Counters in the 9 Clusters 31

CHAPTER 1 - INTRODUCTION

The Pennsylvania Department of Transportation (PENNDOT) has an extensive traffic volume monitoring program aimed at developing factors for the expansion of short-term counts and at monitoring traffic growth throughout the state. A previous project provided a statistical analysis of PENNDOT's existing traffic volume monitoring program to determine if current sample sizes were sufficient, and to provide guidance with respect to short-term initiatives that will make the program more effective and efficient. This research provides a more detailed analysis of the program and investigates alternative ways of developing growth factors. Specifically, three primary objectives were undertaken:

1. To conduct a literature review and national survey of growth factor estimation procedures. Of particular interest were the frequency, duration, and spatial density of counts taken to support the estimation of growth factors.

2. To use historical PENNDOT traffic data to test the impact of reducing the duration of control counts from three weeks. This could lead to the deployment of resources in order to increase the number of such counts.

3. To use historical PENNDOT traffic data to cluster the existing permanent continuously operating automated traffic recorders (ATRs) according to monthly factors and other criteria.

This report is organized so that each objective is presented in a single chapter with a single supporting appendix. Discussion related to the first objective is contained in chapter 2, with supporting materials contained in appendix A. Likewise, the second objective is covered in chapter 3 and appendix B, and the third objective is covered in chapter 4 and appendix C. Each chapter concludes with a summary section that documents the important results and recommendations. Chapter 5 is provided as a consolidation of these summaries.

CHAPTER 2 - LITERATURE REVIEW AND NATIONAL SURVEY

2.0 INTRODUCTION

This chapter documents literature reviewed and a national survey conducted regarding the computation of growth factors from a statewide traffic data collection program. In section 2.1, the results of the literature review are presented and related to the current PENNDOT program. Two main sources are referenced: the current (1995) and proposed (2001) versions of the Federal Highway Administration's (FHWA) Traffic Monitoring Guide (TMG). These two documents provide a good coverage of the topic because (1) the TMG should be used by all states, including PENNDOT, in the conduct of traffic volume monitoring, and (2) the updated TMG (2001) includes the results of valid recent research in its recommendations. Furthermore, a literature search was performed using the search engines EI Compendix, Transportation Research Information Service (TRIS), and National Technical Information Service (NTIS). However, this search revealed no additional pertinent literature since 1990.

Section 2.2 presents the results of a survey of other states' growth factor estimation programs. Similarities and differences with the current PENNDOT program are highlighted.

2.1 Literature Review

2.1.1 Current PENNDOT Program

The database PENNDOT currently uses to calculate growth factors is comprised of 63 permanent continuously operating ATRs and a few hundred short-term control counts. The short-term counts that are used in growth factor estimation are collected three times per year for one week at each site. Each week in the series of three takes place in a different quarter of the year. One count is performed in the quarter of April through June, the second count is collected sometime between July and September, and the last one-week count is collected between October and December. These short-term counts are also used in the Highway Performance Monitoring System (HPMS) program, along with the ATR counts and approximately 5,500 other short-term counts. The other short-term HPMS counts are of 24-hour duration and are taken on a three-year cycle.

For the purposes of growth factor estimation, facilities are grouped into two categories according to functional class: interstates and non-interstates. The state is separated into 11 geographic regions for interstates and 10 geographic regions for non-interstates. For each geographic region - functional class combination, growth factors are computed for urban facilities and rural facilities separately. These differentiations result in 42 different growth factors statewide.

Although it is believed that the current PENNDOT growth factor estimation program produces reasonable results, previous studies have indicated that the sample size may not have been large enough.

2.1.2 Relevant Observations from the Current Traffic Monitoring Guide (1995)

Three primary observations were noted in the 1995 edition of the TMG that are relevant to the current PENNDOT program:

1. Growth factors developed at a specific point, even if from continuous ATR data, are unreliable.

2. For growth factor estimation, the large spatial sample size from the HPMS counts are superior to the large temporal sample size from the continuous ATR program.

3. The primary data used to support the growth factor estimation should be the HPMS counts. A special database collected specifically for growth factor estimation should not be necessary.

These observations are discussed in further detail below.

The following excerpt from the TMG (1995) explains why growth estimation from a point-specific count is questionable, even when the count data are from a fully functional permanent ATR:

> "Growth factors at a point can be best estimated based on the presence of a continuous ATR, assuming that the ATR data is reliable and that the differences found from year-to-year can be attributed to growth. Since it is well known that many extraneous effects contribute to these differences, the assumption may not be realistic in many cases. It should be clear that even with continuous ATRs the site-specific growth factor may be questionable."

This principle points out the weakness of the approach that emphasizes precise estimation of AADT at a limited number of sites. It is then further expanded to describe the type of database that should be used for growth factor estimation. The following is an excerpt from the TMG (1995):

> "The procedure recommended in this Guide (for growth factor estimation) is based on the coverage program because it is believed that the large spatial sample size in the coverage program is superior to the large temporal sample size in the continuous ATR program."

The coverage program referred to is from the HPMS traffic count data collection program. It is important to note that the very basis of this excerpt is that growth factors will be computed using data available from another program, thus implying that a special data collection program is not required. In fact, the discussion of growth factor estimation was a subheading under the larger section dealing with HPMS data collection. Specifically, the following excerpt describes the TMG (1995) recommended data to use in the growth factor estimation computations:

> "Growth factors can <u>also</u> be developed from the coverage program if structured in the manner recommended in this guide (a large number of 48-hour counts taken on a three-year cycle). Since AADT estimates will be developed annually at each HPMS standard sample, the AADT ratios from year-to-year should provide point-specific growth ratios... Since the number of counts is large, system growth estimates developed from the many current counts would benefit by the averaging effect which would be expected to reduce the variability considerably."

2.1.3 Relevant Observations from the Updated Traffic Monitoring Guide (2001)

A draft version of this yet unfinalized document was obtained from the Federal Highway Administration internet site. This version of the TMG (2001) reinforces many of the principles contained in the 1995 version and highlighted in section 2.1.2. They also contain stronger and more specific recommendations for the HPMS traffic counting program, which are summarized below:

1. One third of the HPMS sample/universal sections should be counted each year so that all sections are counted at least every three years.

2. The recommended duration of short-term HPMS counts is 48 hours. The following excerpt explains why:

    > "The most common data collection time periods for traffic volume counts taken with conventional traffic counting equipment are 24- and 48-hour counts. The 48-hour counts are particularly important for the HPMS because common data collection periods for HPMS counts from all states ensures similar levels of accuracy and precision for all volume data in the HPMS database."

The current PENNDOT HPMS data collection program uses 24-hour counts, which are less reliable than 48-hour counts, particularly in rural areas (TMG, 2001). If PENNDOT changes the HPMS data collection program to use 48-hour duration counts to comply with the TMG, this would improve the utility of these counts for use in the growth factor estimation program. Furthermore, it was noted that the PENNDOT Roadway Management System, which is another user of the many short-term counts, is currently being upgraded to accommodate 48-hour counts.

2.1.4 Summary

In summary, both versions of the TMG emphasize the usage of a large spatial sample size in lieu of a large temporal sample size for the growth factor estimation program. Furthermore, both versions recommend the use of HPMS data, supplemented by the continuous ATR data, in the program. They do not indicate that special data collection is required for growth factor estimation, and in fact, it can be deduced through the organization of the report and the context in which the subject is addressed that growth factors will be estimated from data collected for some other purpose.

2.2 National Survey of Growth Factor Estimation Programs

A telephone survey was conducted for as many states as possible to gain insight into their respective growth factor estimation programs. The surveys were more discussion oriented, with emphasis on covering key points, rather than oriented toward a list of specific questions. The following 39 states participated in the survey:

Alabama	Missouri	Virginia
Arkansas	Montana	Washington
Colorado	Nebraska	West Virginia
Delaware	Nevada	Wisconsin
Florida	New Hampshire	Wyoming
Georgia	New Jersey	
Idaho	New York	
Illinois	North Dakota	
Iowa	Oklahoma	
Kansas	Oregon	
Kentucky	Rhode Island	
Louisiana	South Carolina	
Maine	South Dakota	
Maryland	Tennessee	
Massachusetts	Texas	
Michigan	Utah	
Mississippi	Vermont	

The results can be grouped into three main categories:

1. The Data Collection Program
2. The Manner in Which Highways are Grouped
3. The Mechanics of Calculating Growth Factors

Each of these will be discussed separately below. Note that since the primary emphasis of the research is the data collection program, this section is presented in more detail. Less information was gained, and is presented, for the material in the other sections.

2.2.1 The Data Collection Program

The discussion on the data collection program centered around the usage and number of continuous ATRs, and the specifics of the short-term count program.

A. Continuous ATRs

Not surprisingly, each state surveyed reported using its continuous ATRs in the growth factor estimation program. As a benchmarking activity for the continuous ATR program, the number of such stations in each state, along with the total road mileage, are provided in table 2-1.

Table 2-2 provides a rank order list of road mileage per ATR by state. As can be seen, Pennsylvania has a low number of ATRs relative to other states. However, it must be noted that total road mileage is not the only variable that the number of ATRs should be based on. Socioeconomic factors such as population distribution and homogeneity of the state are important considerations in the number of ATRs needed. In addition, the inclusion of non-state owned roads, i.e., county or township roads, in the ATR program is also an important consideration. The average miles per ATR is 944 over the participating states. To achieve a coverage of 944 miles per ATR, Pennsylvania would have to roughly double the number of counters currently in the program.

Table 2-1 - Continuous ATRs By State.

State	# of Stations	Road Mileage*	State	# of Stations	Road Mileage*
Pennsylvania	**63**	**119,281**	Nebraska	65	92,744
Alabama	104	94,228	Nevada	65	35,411
Arkansas	35	95,110	New Hampshire	85	15,124
Colorado	100	85,272	New Jersey	66	35,921
Delaware	60	5,732	New York	107	112,524
Florida	303	115,416	North Dakota	49	86,603
Georgia	134	113,352	Oklahoma	100	112,524
Idaho	162	46,107	Oregon	130	68,481
Illinois	86	137,962	Rhode Island	50	6,050
Iowa	142	112,811	South Carolina	70	64,894
Kansas	102	133,826	South Dakota	51	83,412
Kentucky	65	73,635	Tennessee	32	86,604
Louisiana	60	60,747	Texas	160	296,581
Maine	35	22,639	Utah	81	41,343
Maryland	78	30,189	Vermont	60-70	14,252
Massachusetts	224	35,251	Virginia	270	69,860
Michigan	142	121,482	Washington	140	80,226
Mississippi	45	73,295	West Virginia	52	35,830
Missouri	85	122,847	Wyoming	120	28,458
Montana	73	69,890			

* As of 1998 - Source Highway Statistics 1998, FHWA-PL-99-017, USDOT, Washington, D.C., 1999.

Table 2-2 - Road Mileage per ATR by States in Rank Order.

State	Miles per ATR	State	Miles per ATR
Delaware	96	Michigan	856
Rhode Island	121	Alabama	906
Massachusetts	157	South Carolina	927
New Hampshire	178	Montana	957
Vermont	237	Louisiana	1,012
Wyoming	237	New York	1,052
Virginia	259	Oklahoma	1,125
Idaho	285	Kentucky	1,133
Florida	381	Kansas	1,312
Maryland	387	Nebraska	1,427
Utah	510	Missouri	1,445
Oregon	527	Illinois	1,604
New Jersey	544	Mississippi	1,629
Nevada	545	South Dakota	1,636
Washington	573	North Dakota	1,767
Maine	647	Texas	1,854
West Virginia	689	Pennsylvania	1,893
Iowa	794	Tennessee	2,706
Georgia	846	Arkansas	2,717
Colorado	853		

B. Short-Term Count Program

Most states operate an extensive short-term count program with sites numbering in thousands and even tens of thousands. As such, most states are focusing their data collection efforts on assembling a large spatial database. One notable exception is Delaware, which relies entirely on its 60 continuous ATR sites for growth factor estimation. The details of the short-term data collection program for each state are provided in appendix A. Tables 2-3, 2-4, and 2-5 document the number of states that fall into certain categories based on number of counts (table 2-3), duration of counts (table 2-4), and counting frequency (table 2-5). In each table, Pennsylvania is not included in the number of responding states, but the category in which it would fall is noted. Finally, note that not all participating states responded to all questions.

Table 2-3 - Number of Short-Term Count Sites.

Number of Counts	Number of States	
>10,000	8	
5,000-10,000	11	
2,500 - 5,000	7	
1,000 - 2,500	5	
<1,000	3	<--Pennsylvania

Table 2-4 - Duration of Short-Term Counts.

Count Duration	Number of States	
24 hours	8	
48 hours	14	
Greater than 48 hours	6*	<--Pennsylvania
Combination of 24 and 48	5	
Other Combinations	3	

*Washington and New York use 72-hour durations. Montana uses 36 hours. Nevada, Vermont, and New Hampshire count for one week. No states indicated using durations longer than one week.

Table 2-5 - Count Frequency.

Count Frequency	Number of States
Every Year	3
Every Other Year	4
Every Three Years	20
Greater Than Three Years	2
Some Every Year and Some at a Greater Frequency	4
Other Combinations	4

<--Pennsylvania

Given that the current TMG (1995) recommends a large number of 48-hour counts taken on a three-year cycle, it appears from the data that most states are using the HPMS data for growth factor estimation.

2.2.2 The Manner in Which Highways are Grouped

It was expected that most states, like Pennsylvania, group the facilities by some combination of functional class, area type, and geography. The following are the key points relative to highway grouping:

> Nevada does not perform any grouping of highway facilities. Growth is predicted for specific points as needed, and sometimes for the state as a whole.
>
> Massachusetts is the only state surveyed that does not distinguish between urban and rural facilities.
>
> Wyoming, Montana, Nebraska, Nevada, and Rhode Island do not have divisions within the state for growth prediction. In Texas, each highway segment is counted each year, so there is no need for aggregation to estimate growth for uncounted facilities. However, they aggregate the entire state to estimate statewide vehicle miles traveled (VMT) growth.
>
> In general, many states use finer geographic divisions of their state than Pennsylvania. New Hampshire and Vermont use a town-based system. Thirteen (13) other states reported predicting growth factors at the county level.

All states except Nevada grouped the highways by functional class in some manner. While there were many different variations, many followed the interstate - non-interstate grouping that Pennsylvania uses. Others included a more detailed breakdown of the non-interstate grouping, and a National Highway System (NHS) - Non-NHS groupings.

One grouping mechanism reported by several states was seasonality. Other grouping mechanisms that were reported once each were truck percentage and road ownership.

2.2.3 The Mechanics of Calculating Growth Factors

Lastly, states were surveyed to determine the way in which the growth factors are calculated. Key points determined were whether the growth rates were weighted towards higher volume facilities, the number of years of data considered in calculation of the growth rate, and whether socioeconomic factors, such as employment and population, were factored into the calculation.

The number of states reporting that growth factors are weighted towards higher volume facilities were minimal and included only Illinois, Louisiana, Mississippi, Vermont, and Wyoming. This is an indication that most states perform an average of the growth factor at each site rather than calculating the growth factor of the combined total of all AADTs.

Several states reported using socioeconomic factors in some way in the calculation of growth factors. It is recognized that any state estimating growth from urban or statewide traffic assignment models is using socioeconomic factors. States using socioeconomic factors as part of trend-analysis type growth forecasting include: West Virginia, Oklahoma, Wyoming, South Dakota, Nevada, New Jersey, and Colorado. Population was used by four of the states and employment was used by two states. Tennessee, Kentucky, and Virginia reported using land use / land use changes as part of the forecast and / or in data collection planning.

Finally, it was found that most states use more than two years of data to project growth, as they are performing linear regression on a historically longer database. Table 2-6 provides a breakdown of the number of years considered in growth rate calculations by the number of states employing them.

Table 2-6 - Number of Years Considered in Growth Factor Estimation.

Years Considered	Number of States	
2	2	<--Pennsylvania
3 - 5	6	
6 - 10	6	
11 - 15	4	
16 - 20	16	
> 20	2	

2.3 SUMMARY

The literature review and national survey provided significant insight into alternatives or directions that might be taken to improve the current PENNDOT growth factor estimation program. Both the literature review and national survey indicated that the number of sites in the short-term count program needs to be greatly expanded, while the three-week count duration can be shortened to less than one week, and likely to 48 hours. One way to assemble the required data for growth factor estimation without developing a new data collection program is to begin using longer duration (48-hour) counts in the HPMS program so that these data can also be used for growth factor estimation. The statistical research presented in chapter 3 investigated the advantages and disadvantages of a large temporal sample size versus a large spatial sample size to give clearer answers to questions related to count duration, frequency, and number of sites.

Also, note that the number of growth factors (42) calculated by PENNDOT was not found to be unusually large, and in fact was less than most states. This has two important ramifications for the current program. First, suggestions that sample sizes in the current PENNDOT program are deficient because the program simply has too many categories should be dismissed. Second, a larger spatial sample size provides opportunities for expansion of the number of growth factors computed, if such an expansion is desired.

CHAPTER 3 - TESTING OF SPATIAL VERSUS TEMPORAL COUNTING PROGRAMS

3.0 INTRODUCTION

The purpose of this chapter is to discuss the research performed on the process of growth factor estimation, and make recommendations to guide the future development of the program in PENNDOT. The research performed focused on two areas: the accuracy gains resulting from increasing spatial sample size, and the losses resulting from decreasing temporal sample size.

As a review, the current PENNDOT growth factor estimation process includes a relatively small spatial sample size with a long duration (large temporal sample size) at each count station. Throughout the state, less than 250 count stations are used to produce 42 different growth rates. At each count station, the duration of the counts is either 365 days for the 63 ATRs or 21 days for the remaining stations, called "control counts." Previous statistical testing indicated sample size deficiencies with the growth factors this process produced. The literature review and national survey suggested that the process might be improved by using a larger spatial sample size with shorter count durations.

The usage of the 63 ATRs, which collect data continuously for 365 days per year, is not in question. These counters have broader uses in addition to providing data for the estimation of growth factors. The primary issue addressed in this report is the value--to the growth rate estimation process--of the data from the three-week control counts relative to the 24-hour HPMS counts, which are currently collected by PENNDOT but not used in growth factor estimation. These 24-hour HPMS counts are much more numerous and provide superior spatial coverage of the state.

3.1 Methodology

Two studies were performed during this portion of the project. The first was conducted in Allegheny County, and involved a comparison of the growth rate in that county's four ATRs to both the growth rate in the control count data and the HPMS data. It was postulated that the benefits of increasing the spatial sample size could be determined by comparing how closely each short-term count database tracked the patterns exhibited at the ATRs. In addition, since several years of data were available in the HPMS database, the growth rates were computed in two ways: using only the last two available years, and using regression analysis on the entire database. In this way, the benefits of using a historical database can be determined.

The second study was conducted with the entire statewide control count database. Each control count station has up to 21 days of data per year. From these 21 days, shorter durations of one day, two days, three days, and one week were randomly extracted, and the growth rates were computed. These growth rates were then assembled by group and compared to the growth rates at each ATR in their respective group. The goal was to see how much the error in projected growth rate increased with decreasing count duration.

The details of each study methodology are presented below.

3.1.1 HPMS vs Control Count Comparison

1. The ATRs that are in Allegheny County were identified. For each ATR, the 1998, 1999, and 2000 AADT were used to compute the average annual growth rate from 1998 to 2000. A growth rate was computed for both 1998 to 1999 and 1999 to 2000. These two growth rates were then averaged together to compute the 1998 to 2000 annual growth rate. The difference between the annual growth rate computed in this manner and the annual (compound) growth rate computed using regression analysis is negligible when the growth is small and the time horizon is short. Using the three-year period produced more stable growth rates than using the last two years only.

2. The control count stations included in the groups that contain Allegheny County were identified. Note that only the urban control counts were used since all the ATRs and HPMS counts were urban. At each control count station, the data available in 1998, 1999, and 2000 were used to compute an annual growth rate. These growth rates were then aggregated (both weighted by AADT and unweighted) by functional class (interstate / non-interstate) groupings. These growth rates were then compared to the ATR growth rate in the same interstate / non-interstate grouping. The comparison was conducted by averaging the absolute value of the differences in growth rates across the four ATRs.

3. The process described in (2) was repeated using the HPMS data for Allegheny County. The last two years of data available were used to estimate an annual growth rate at each site for years 1999 and 2000. Average growth rates were then calculated for 1999 and 2000. The process was then modified to include all available years of data in the computation of the growth rate. Regression analysis was used assuming annual compounding in the growth rate computation.

4. The averages of absolute errors for the above procedures were then compared. The procedure yielding the lowest value is deemed to have performed the best.

3.1.2 Count Duration Comparison

1. The growth rates from 1999 to 2000 were identified for each of the 63 ATRs. These growth rates were assembled according to the 42 growth factor estimation groups (groups with a sample size of zero were omitted), however, they were not aggregated.

2. At each control count station, a count from each year was randomly extracted for each of the following durations:
   One day
   Two consecutive days
   Three consecutive days
   One week - Note that in some cases, the week that was randomly selected did not have seven full days

3. A growth rate was calculated at each station using the above-mentioned randomly extracted counts, as well as the entire (three-week) database.

4. The growth rates at each control count station were then compared to the growth rates at each ATR in the group to which they belonged. The comparison was conducted by averaging the absolute value of the differences across the ATRs in their group. A final comparison was then made by averaging all the absolute errors across all groups, yielding a single value for each count duration.

Details related to the "HPMS vs Control Count" research can be found in appendix B. The details related to the "Control Count Duration" research are too voluminous to merit inclusion in the report, but are provided in electronic (spreadsheet) form. The results of each are discussed below.

3.2 Results

3.2.1 HPMS vs Control Count Comparison

There are four ATRs in Allegheny County: two interstate and two non-interstate. Details on each ATR are provided in table 3-1.

Table 3-1 - Allegheny County ATR Information.

ATR #	Group	2000 AADT	Average Growth Rate ('98-'00)
203	Urban Non-Interstate	21,023	0.90%
375	Urban Non-Interstate	23,990	-0.55%
208	Urban Interstate	65,332	0.05%
309	Urban Interstate	106,761	-0.10%

The sample sizes and growth rates based on the control count data are shown in table 3-2.

Table 3-2 - Control Count Information.

Group	Sample Size*	Average Growth Rate ('98-'00)
Urban Non-Interstate	16 - 17	1.15%
Urban Interstate	2 - 4	3.05%

*The first number represents the sample size for computing a growth rate between '98 and '99, the second number is the sample size for computing the growth rate between '99 and '00.

The sample sizes and growth rates based on the HPMS data are shown in table 3-3. Four methods were used to compute the growth rate. The first two methods use only data from the last two available years, but differ in that one is a weighted average (by AADT) in the aggregation. The last two methods differ from the first two in that all years of data available were used to estimate the growth factor.

Table 3-3 - Allegheny County HPMS Growth Rate Information.

Group	Sample Size	Last Two Available Years		All Years	
		Unweighted	Weighted	Unweighted	Weighted
Urban Non-Interstate	56 - 65	1.99%	2.56%	1.15%	2.18%
Urban Interstate	2 - 4	1.74%	0.10%	-0.66%	-2.77%

The growth rates in tables 3-2 and 3-3 were then compared to the growth rates of the ATRs as shown in table 3-1. In each case, the absolute value of the difference in growth rates was averaged across the ATRs. For example, table 3-4 shows the calculations for Average Absolute Error of the Control Count Growth Estimation.

Table 3-4 - Absolute Error of the Control Count Growth Estimation.

ATR	ATR Growth	Control Count Growth	Absolute Error
203	0.90%	1.15%	0.25%
375	-0.55%	1.15%	1.70%
208	0.05%	3.05%	3.00%
309	-0.10%	3.05%	3.15%
		AVERAGE	**2.03%**

This analysis indicates that on average, the growth rate predicted at the ATRs by the control count data was in error by 2.03 percent. Note that this is an absolute error, although it is expressed as a percentage. Since growth rates are typically on the order of 2 percent, this could also be interpreted as indicating that the control count data could yield a relative error of approximately 100 percent or more.

The Average Absolute Errors for each of the methods tested are shown in table 3-5. As can be seen, usage of all years of HPMS data (unweighted aggregation) was the best performing method. This method was superior to the others by a relatively large margin. Compared to the Control Count Data, this method produced the same growth rate for the non-interstates, and a much closer estimate of the growth rate exhibited at the interstate ATRs. A review of the non-interstate growth rate in light of the sample sizes demonstrates that the 16 control counts yielded the same growth rate as the 65 HPMS counts. A similar review of the interstates shows that increasing the spatial sample size from 2 control counts to 4 HPMS counts significantly improved the quality of the projection, even though the temporal sample size was reduced from 42 days +/- to 4 days. This suggests that increasing spatial sample size is highly beneficially, particularly when the sample size is small. In general, note that all but one of the alternatives that utilized the 24-hour HPMS data outperformed the three-week Control Count Data alternative.

These results are a confirmation of the findings of the literature review and national survey. It suggests that PENNDOT should consider using a larger spatial database, such as might be provided by the full HPMS databases. As a tradeoff, they may consider less data collection than the three one-week counts at the control count stations. The increased errors expected from shortening the data collection duration are discussed in the following section.

Table 3-5 - Comparison of Average Absolute Errors of the Various Methods.

Method	Average Absolute Error
HPMS Data, all years included	0.80%
Weighted Average of HPMS Data, last two available years	1.26%
HPMS Data, last two available years	1.79%
Control Count Data	2.03%
Weighted Average of HPMS Data, all years included	2.37%

3.2.2 Count Duration Comparison

The "Average Absolute Error" comparison method, as discussed above, was also used in this analysis. Using the control count data for the entire state, different duration counts were extracted at each station and growth factors computed. These growth factors were then assembled by group and compared to each ATR in the group using the "Average Absolute Error" criterion detailed above. These were aggregated to the facility-type level by averaging across all the groups within a certain facility type. These results are shown in table 3-6. They were then aggregated across all facility types, as shown by the bottom line in table 3-6. The five durations compared in this analysis are one day, two days, three days, one week, and three weeks (existing program).

Table 3-6 - Results of Count Duration Comparison.

Facility Type	Average Absolute Error				
	1 Day	2 Days	3 Days	1 Week	3 Weeks
Rural Non-Interstate	2.98	5.88	4.79	3.46	6.81
Urban Non-Interstate	2.31	2.84	1.66	1.66	2.15
Rural Interstate	2.86	1.32	1.71	3.38	4.64
Urban Interstate	4.53	4.22	2.45	2.38	2.63
AVERAGE	**3.02**	**3.98**	**2.93**	**2.66**	**4.23**

The best performing duration overall was one week, followed by three days, then one day, two days, and finally three weeks. There is a general trend suggesting that as count duration is increased, the error is decreased, which was expected. However, the two-day and three-week count durations do not follow this trend. The three-week count duration was the worst performing method overall. In no case was it the best performing method for a facility type, as it scored no better than third-best in any case. The poor performance of the three-week period was unexpected, and may be related to the increased probability of non-typical days being included in the database as more days are included.

At worst, this analysis indicates that a significant difference could not be detected between the durations. At best, it indicates that a one-week duration should be used at the control count stations. In all likelihood, due to limitations of the sample size, it should be interpreted as indicating that three-week durations provide little benefit over shorter durations of a week or less.

3.3 Conclusions and Recommendations for the Growth Factor Estimation Process

Based on the results of the nationwide survey, literature review, and the results of this testing, it is recommended that PENNDOT (1) include the 24-hour HPMS counts in the growth factor estimation process, (2) shorten the duration of the control counts to one week or less, (3) use more than the last two years in growth rate computations, and (4) always manually inspect and clean the data prior to inclusion in the analysis.

As stated in the literature review, the new TMG strongly urges states to collect 48-hour HPMS counts to provide uniformity among the states. Because of their large number, the HPMS database will improve the quality of the growth factor estimation process, even if they are only 24-hour counts, but particularly so if they are 48-hour counts. Since they are short, particular care should be given in the selection of the day(s) that the counts are collected. Perhaps a review of the ADT to AADT adjustment factors (and/or their standard deviations) should be made to select days that have historically produced stable counts.

The duration of the control counts can be reduced. A maximum duration of one week is recommended since the same effort is involved in collecting a single count at these stations, regardless of duration. However, durations shorter than one week could also be used. The results of the analysis indicate that a significant drop-off in quality will not occur, and the effort of placing the counter in the field for the second and third week counts can be saved.

A historical database should be used in the estimation of growth factors. In this research, data up to eight years of age were used. This provided an advantage over using only the last two available years. The analysis did not address the number of years to use, however, that is likely not a significant item at this point, since most of the data available for growth factor estimation will only be a few years old. As a larger historical database is assembled, this issue should be revisited. Note that in the nationwide survey, of the 36 states responding to the question of "years" considered, the most popular answer was 16 to 20 years, as 16 states utilized this duration.

Regardless of whether the above-mentioned modifications are made, it is strongly recommended that the count data be rigorously inspected and cleaned before use. Significant effort was put into cleaning the data before the testing was performed. Had the data not been cleaned, the results would have been significantly affected. The same could be said for the computation of the growth rates. In addition, it was noted that the control counts rarely had three full weeks in both directions at a particular station. This likely has an affect on the quality of the growth factors that use these counts. It is also another reason to limit the control count duration to a single week or less, since the current effort rarely produces the desired counts.

3.4 Limitations of the Research

The "HPMS vs. Control Count Comparison" research was limited by the number of ATRs in Allegheny County. There were only four ATRs available for comparison purposes. More ATRs would have provided a more reliable comparison. In addition, there were no rural facilities in Allegheny County. The results are entirely based upon urban facilities.

The "Count Duration Comparison" research was limited by the number of ATRs in each of the 42 groups. Most groups had three or fewer ATRs.

3.5 Summary of Recommended Modifications to the Growth Factor Estimation Process

It is recommended that PENNDOT use all of the short-term HPMS counts in the growth factor estimation process. The duration of these counts should be a minimum of 48 hours as per the updated TMG.

No more than one week of data is needed at the Control Count Stations

Historical (more than last two years) data should be used in estimating the growth factors

The raw data need to be rigorously inspected and cleaned before use.

CHAPTER 4 - CLUSTER ANALYSIS OF ATRS

4.0 INTRODUCTION

The purpose of this chapter is to document the cluster analysis performed on the year 2000 ATR count data. The purpose of the clustering was two-fold. First, the TMG requires the ATRs to be clustered according to seasonal variations, as measured by the monthly adjustment factors (ratio of annual ADT to monthly ADT). Second, from the cluster analysis results, it can be determined whether the existing system of Traffic Pattern Groups (TPGs) is adequate.

This chapter is organized in the following way. In section 4.1, details related to the clustering analysis are presented. This section provides a description of the technical aspects of this research. In addition, it provides, along with appendix C, the necessary information so that this work can be repeated in the future, if desired. Section 4.2 provides conclusions and recommendations, including a precision analysis of the TPGs / clusters.

4.1 Cluster Analysis

This section describes in detail the work performed in the cluster analysis. When used with the supplementary information contained in appendix C, it also provides a map for those desiring to repeat the analysis in the future.

In general, the following tasks were performed in the cluster analysis:

1. The recreational ATRs were identified and eliminated from the cluster analysis. In addition, due to discrepancies in the functional class of counter 383, it was also excluded from clustering.

2. A cluster analysis was performed solely based on the seasonality of the ATRs, as measured by the monthly factors. This a very interesting analysis in that it clusters the ATRs without any other constraints and shows which have the closest matching seasonal variations. It lacks practicality however, as facilities from different functional classes are clustered together, making it nearly impossible to tell to which cluster a count station (e.g., a short-term count) from outside the 63 ATRs might belong.

3. The cluster analysis was repeated while forcing facilities of the same functional class together. This yielded practical results that have the same usefulness as the current traffic pattern groups.

4. Finally, current traffic pattern groups 6, 7, 8, and 9 distinguish between "north rural" and "central rural" facilities. This grouping was put to the test in another

cluster analysis to determine if it is consistent with the seasonal patterns at the ATRs in these groups.

4.1.1 Identification and Elimination of Recreational ATRs

As described in the TMG, a high value of the coefficient of variation (MCV) for the monthly factors is an indication of high seasonality, which is an attribute associated with "recreational" roads. MCV is calculated as the ratio of the standard deviation of the monthly factors to their mean (x 100). The year 2000 traffic data from the ATRs were first sorted based on the values of MCV to examine the patterns and decide on the recreational group. It was decided that the ATRs with MCV values of more than 20 percent for rural roads and more than 12 percent for urban roads would be categorized as "recreational." Appendix C shows the monthly factors, means, standard deviations, and MCVs for each of the 63 ATRs. Based on an examination of this criterion, the following ATRs were identified as recreational and were not included in the cluster analysis:

27 - SR 66 - Elk County
306 - SR 507 - Pike County
371 - I-70 - Fulton County
384 - SR 4022 - Tioga County

Also, counter 383 (SR 150 - Clinton County) was listed as an urban local road in the traffic data, but is listed under TPG 8 (North Rural Collector) in the PENNDOT literature. Due to this discrepancy, it was excluded from the cluster analysis.

4.1.2 Cluster Analysis Unconstrained by Functional Class

As prescribed in the TMG, the monthly factors (ratios of AADT to MADT) were used to measure the similarity between ATRs and to group them into clusters. Usually, the data need to be "standardized" prior to the cluster analysis. Data are standardized by subtracting from each value in the set the mean of that set and dividing this result by the standard deviation. Standardized data have a mean of zero and a standard deviation of 1. The main benefit of standardizing the data is that all values in the set then have the same weight. However, since the monthly factors are ratios that are all close to the value of 1.0, there was no need to standardize the data.

Different clustering methods are available for the grouping process. As suggested in the TMG, the "Ward's Minimum Variance" method was used to perform the cluster analysis. In this method, the distance between two clusters is calculated as the sum of squares between the two clusters added up over all the variables. The method starts with all observations considered as individual clusters. At each iteration, the within-cluster sum of squares is minimized over all partitions obtainable by merging two clusters from the previous iteration. The merging process continues until all observations are merged into one cluster. Ward□s method tends to join clusters that have a small number of observations and, hence, tends to produce clusters with roughly similar number of observations. Judgement must be used to determine at what level the cluster analysis has achieved a sufficient level of aggregation.

A routine was written for use in the SAS (Statistical Analysis System) program to perform the cluster analysis. It is provided in appendix C along with the raw results of the clustering. From these results, the ATRs were grouped into nine clusters as shown in table 4-1. This table shows the process of how groups were formed in the cluster analysis. The earliest point in the process is at the bottom of the table, and the process of clustering continues as one moves up the table. Entries in the table show the number of the group formed followed by the number of ATRs in the group. The group number corresponds to the order followed in the clustering program. For example, in the last step, group number 1 was formed with all 58 ATRs. In the previous step, there were two groups (2 and 3). The last entries in each column (bottom of the table) show the cluster number (e.g. CL 1), and the numbers of the ATRs included in the cluster. Table 4-2 provides a listing of the ATRs in each of these clusters. Not counting the recreational group, these would be the traffic pattern groups based on this analysis. However, as discussed earlier, these clusters are not suitable as traffic pattern groups because functional classes are mixed among the clusters. This does, however, demonstrate which ATRs have the most similar seasonal patterns.

Table 4-1 - Clustering Process - Unconstrained by Functional Class.

<table>
<tr><td colspan="11">GP 1 (58)</td></tr>
<tr><td colspan="6">GP 2 (29)</td><td colspan="5">GP 3 (29)</td></tr>
<tr><td colspan="5">GP 4 (25)</td><td>GP 8 (4)</td><td colspan="4">GP 6 (24)</td><td>GP 9 (5)</td></tr>
<tr><td>GP 10 (9)</td><td colspan="3">GP 11 (15)</td><td>392</td><td>CL 5</td><td colspan="3">GP 7 (15)</td><td>GP 16 (9)</td><td></td></tr>
<tr><td>CL 1</td><td colspan="2">GP 14 (12)</td><td>GP 18 (3)</td><td></td><td>2, 207, 216, 374</td><td>19</td><td>GP 22 (8)</td><td>GP 20 (6)</td><td>CL 8</td><td>CL 9</td></tr>
<tr><td>4, 5, 24 29, 48, 364, 372, 385, 393</td><td>GP 24 (5)</td><td>GP 33 (7)</td><td></td><td></td><td></td><td></td><td>CL 6</td><td>CL 7</td><td>203, 205, 208, 210, 349, 370, 375, 379, 390</td><td>8, 206, 309, 330, 377</td></tr>
<tr><td></td><td>CL 2</td><td>CL 3</td><td>CL 4</td><td></td><td></td><td></td><td>15, 18, 20, 40, 362, 380, 381, 382</td><td>301, 326, 334, 378, 386, 389</td><td></td><td></td></tr>
<tr><td></td><td>3, 323, 376, 391, 394</td><td>51, 304, 328, 360, 367, 373, 388</td><td>1, 363, 387</td><td></td><td></td><td colspan="3"></td><td></td><td></td></tr>
</table>

Table 4-2 - Clusters Unconstrained by Functional Class.

Cluster Number	ATR Numbers
1	4, 5, 24, 29, 48, 364, 372, 385, 393
2	3, 323, 376, 391
3	51, 304, 328, 360, 367, 373, 388
4	1, 363, 387
5	2, 207, 216, 374
6	15, 18, 20, 40, 362, 380, 381, 382
7	301, 326, 334, 378, 386, 389
8	203, 205, 208, 210, 349, 370, 375, 379, 390
9	8, 206, 309, 330, 377

As noted, one major problem with the clustering scheme in tables 4-1 and 4-2 is that in a number of cases, roads that have the same functional classification are grouped in different clusters. This would create a problem when the seasonal factors calculated for a cluster are applied to a facility from outside these ATRs (e.g., the location of a short-term count). It is desirable to have uniformity within each of the clusters with respect to functional classification, so that the factors calculated for the cluster may be applied to any new road that matches any of the functional classifications of the ATRs in the cluster. One approach to this is to use judgement to move some of the ATRs shown in Table 4-2 such that, as much as possible, ATRs with the same functional classification number belong to the same cluster. This approach was followed in the example given in the TMG. Following this approach, the ATRs were grouped in seven clusters as shown in table 4-3.

Table 4-3 - Results Based on Unconstrained Clustering and Arbitrary Adjustment.

Cluster	Functional Class	ATR Numbers
Rural Interstates	1	207, 216, 370, 372, 374, 392, 393
Rural Other Principal Arterials	2	4, 19, 24, 323, 326, 334, 349, 360, 363, 378
Rural Minor Arterials	6	1, 2, 3, 15, 40, 48, 51, 328, 367,390, 391
Rural Major and Minor Collectors	7 & 8	5, 29, 362, 364, 385, 386, 388, 389
Urban Interstates	11	205, 208, 210, 309, 373, 376, 377, 394
Urban Other Principal Arterials	12 & 14	8, 203, 206, 301, 304, 330, 375
Urban Minor Arterials and Collectors	16 & 17	18, 20, 379, 380, 381, 382

An alternative approach would be to use the functional classification codes as part of the clustering procedure. This is described below.

4.1.3 Clustering Process that Includes Functional Class

To follow this approach a new variable, "functional code value" was introduced to assign a weight to each code as shown in table 4-4.

Table 4-4 - Values Assigned to Functional Classification Codes.

Functional Class No.	Functional Class Description	Value
1	Rural - Principal Arterial - Interstate	10
2	Rural - Principal Arterial - Other	20
6	Rural - Minor Arterial	30
7	Rural - Major Collector	40
8	Rural - Minor Collector	50
9	Rural - Local System	60
11	Urban - Principal Arterial - Interstate	110
12	Urban - Principal Arterial - Other Freeways or Expressways	120
14	Urban - Principal Arterial - Other	130
16	Urban - Minor Arterial - Other	140
17	Urban - Collector	150
19	Urban - Local System	160

With the inclusion of the new variable, the functional classification will carry more weight in the cluster analysis than any of the monthly factors. In addition, ATRs for rural roads will tend to be clustered with those of other rural roads with similar or close functional code values. The same is true for ATRs of urban roads. As an example, rural-principal arterial-interstate roads will have a better chance to be grouped with rural-principal arterial-other roads than with rural-minor-collector roads or with any urban road.

A SAS routine was written to perform the new cluster analysis and is provided in appendix C along with the raw results. A summary of the clustering process is provided in table 4-5 and a summary of the results is presented in table 4-6. Note that table 4-5 follows the same format as table 4-1. Furthermore, table 4-6 contains a potential set of traffic pattern groups, excluding the recreational group.

Table 4-5 - Clustering Process with Functional Class Used as a Parameter.

GP 1 (58)						
GP 2 (37)				GP 3 (21)		
GP 5 (20)		GP 6 (17)		GP 4 (15)		GP 9 (6)
GP 7 (9)	GP 15 (11)	GP 10 (7)	GP 12 (10)	GP 8 (7)	GP 11 (8)	**CL 7**
CL 1	**CL 2**	**CL 3**	**CL 4**	**CL 5**	**CL 6**	
5	1	207	4	8	205	18
29	2	216	19	203	208	20
362	3	370	24	206	210	379
364	15	372	323	301	309	380
385	40	371	326	304	373	381
386	48	372	334	330	376	382
387	51	374	349	375	377	
388	328	392	360		394	
389	367	393	363			
	390		378			
	391					

Table 4-6 - Clusters Constrained by Functional Class.

Functional Class	Cluster Number (From Table 5)	Group Name	ATR Numbers
1	CL 3	Rural Interstates	207, 216, 370, 372, 371, 372, 374, 392, 393
2	CL 4	Rural Other Principal Arterial	4, 19, 24, 323, 326, 334, 349, 360, 363, 378
6	CL 2	Rural Minor Arterial	1, 2, 3, 15, 40, 48, 51, 328, 367, 390, 391
7 & 8	CL 1	Rural Collectors	5, 29, 362, 364, 385, 386, 387, 388, 389
11	CL 6	Urban Interstates	205, 208, 210, 309, 373, 376, 377, 394
12 & 14	CL 5	Urban Other Principal Arterial	8, 203, 206, 301, 304, 330, 375
16 & 17	CL 7	Urban Others	18, 20, 379, 380, 381, 382

As can be seen, this matches the results obtained when the unconstrained clustering was performed and the ATRs were then arbitrarily adjusted based on functional class. It also is very similar to the existing set of traffic pattern groups used by PENNDOT. The nine non-recreational traffic pattern groups used by PENNDOT are shown in table 4-7 along with the cluster to which they match.

Table 4-7 - Comparison of Clusters and Existing Traffic Pattern Groups.

Traffic Pattern Group (TPG)		Cluster No (From Table 4-5)	Comments
No.	Name		
1	Urban Interstate	CL 6	exact match
2	Rural Interstate	CL 3	exact match except for 371, which was excluded
3	Urban Principal Arterial	CL 5	exact match
4	Rural Principal Arterial	CL 4	exact match
5	Urban Minor Art or Coll	CL 7	exact match
6	North Rural Minor Arterial	CL 2	27 was excluded, north and central areas not separated
7	Central Rural Minor Arterial	CL 2	
8	North Rural Collector	CL 1	383 & 384 excluded, north and central areas not separated
9	Central Rural Collector	CL 1	

The only difference between the results of the cluster analysis and the existing set of traffic pattern groups is that the cluster analysis did not yield rural facilities in the "north" and "central" portions of the state as separate categories from one another. Before concluding that there is no difference, another cluster analysis was performed using just the rural minor arterials and collectors (Functional Classes 6, 7, and 8). When these facilities are clustered, if the north and central facilities cluster into separate groups, then the existing PENNDOT traffic pattern groups should be maintained.

4.1.4 Clustering of Rural Minor Arterials and Collectors

In this analysis, three cluster analyses were performed; one used just the minor arterials, one used just the collectors, and a third used all of them. The SAS routine and raw SAS results are provided in appendix C. The clustering process for the minor arterials is provided in table 4-8. For each ATR, it is noted whether the facility is considered "north" (N) or "central" (C).

Table 4-8 - Rural Minor Arterial Clustering Process.

<table>
<tr><td colspan="4">GP 1 (11)</td></tr>
<tr><td colspan="3">GP 2 (8)</td><td>GP 7 (3)</td></tr>
<tr><td colspan="2">GP 3 (7)</td><td rowspan="3">2 (N)</td><td rowspan="3">15 (C), 40 (C), 390 (C)</td></tr>
<tr><td>GP 4 (3)</td><td>GP 5 (4)</td></tr>
<tr><td>3 (N), 391 (C), 48 (N)</td><td>1 (C), 51 (N), 328 (N), 367 (C)</td></tr>
</table>

If these facilities are to be clustered into two groups, then the results at the second line from the top in Table 4-8 are pertinent. At this point, GP 2 contains 8 ATRs and GP 7 contains 3. Of the 8 ATRs in GP2, the majority are "north" minor arterials. All of the ATRs in GP 7 are "central" minor arterials. Therefore, the facilities were roughly divided according to the north / central division.

The clustering process for the collectors is provided in table 4-9. As can be seen, when they are clustered into two groups, GP 2 consists of entirely "central" facilities, while GP 3 consists of 3 "northern" facilities and only one "central" facility. In this analysis, the facilities were strongly divided according to the north / central division.

Table 4-9 - Rural Collector Clustering Process.

<table>
<tr><td colspan="3">GP 1 (9)</td></tr>
<tr><td colspan="2">GP 2 (5)</td><td>GP 3 (4)</td></tr>
<tr><td>GP 4 (3)</td><td>GP 7 (2)</td><td rowspan="2">5 (N), 385 (N), 364 (C), 29 (N)</td></tr>
<tr><td>386 (C), 389 (C), 387 (C)</td><td>362 (C), 388 (C)</td></tr>
</table>

Finally, all of the rural minor arterials and collectors were used in a single analysis. The results are provided in table 4-10. Like tables 4-8 and 4-9, (N) and (C) are used to note whether the facility is "north" or "central." In addition, inside the parenthesis, the functional class code of the facility is noted (6 = minor arterial, 7 or 8 = collector).

In this analysis, three of the four existing PENNDOT traffic pattern groups in question emerged as groups. TPG 6 - North Rural Minor Collectors emerged as a sub-group of GP7, TPG 8 - North Rural Collectors emerged as GP 8, and TPG 7 - Central Rural Minor Collectors emerged as GP 10. TPG 9 - Central Rural Collectors did not emerge as a group because there were only two counters in this group and they were scattered among the other groups.

Table 4-10 - Combined Rural Minor Arterial and Collector Clustering Process.

<table>
<tr><td colspan="9">GP 1 (20)</td></tr>
<tr><td colspan="5">GP 2 (13)</td><td colspan="4">GP 4 (7)</td></tr>
<tr><td colspan="4">GP 3 (10)</td><td>GP 8 (3)</td><td colspan="3">GP 5 (6)</td><td rowspan="3">2
(N6)</td></tr>
<tr><td colspan="2">GP 7 (6)</td><td colspan="2">GP 10 (4)</td><td rowspan="2">386(N7)
387(N8)
389(N7)</td><td colspan="2">GP 6 (5)</td><td rowspan="2">29
(N7)</td></tr>
<tr><td>1(C6)
391(C6)</td><td>51(N6)
328(N6)
388(C7)
367(N6)</td><td>15 (C6)
50 (C6)</td><td>362 (C6)
390 (C6)</td><td>3(N6)
5(N7)
385(N8)</td><td>48(N6)
364(C7)</td></tr>
</table>

In conclusion, it is recommended that the existing set of traffic pattern groups, which differentiates between "north"and "central" for rural minor arterials and collectors, be retained.

4.2 Conclusions and Recommendations

In compliance with the TMG, a cluster analysis of the PENNDOT ATRs was performed. Two key findings were made relative to the existing set of traffic pattern groups.

The existing set of traffic pattern groups should be retained.

Four counters (27, 306, 371, and 384) were identified as having some of the attributes of "recreational" facilities. Counter 306 is considered recreational in the existing set of traffic pattern groups. The others may be considered for reclassification as recreational.

In closing, a precision analysis was performed for the nine non-recreational traffic pattern groups (excluding the ATRs identified as recreational and ATR 383 which had discrepancies). The results are shown in table 4-11. Precision levels less than 10 percent are considered acceptable according to the TMG. As can be seen, all TPGs are acceptable.

Table 4-11 - Precision Analysis for Counters in the 9 Clusters.

TPG (Cluster)	Sample Size	Precision (%)
1 - Urban Interstate	7	3.53
2 - Rural Interstate	8	4.78
3 - Urban Principal Arterial	7	3.48
4 - Rural Principal Arterial	10	2.58
5 - Urban Minor Arterial or Collector	6	2.57
6 - North Rural Minor Arterial	6	5.00
7 - Central Rural Minor Arterial	6	2.94
8 - North Rural Collector	3	5.59
9 - Central Rural Collector	6	3.53

CHAPTER 5 - SUMMARY OF CONCLUSIONS AND RECOMMENDATIONS

5.0 INTRODUCTION

At the end of chapters 2, 3, and 4, the key findings and recommendations of the research detailed in each respective chapter are provided. These are consolidated in this chapter to provide a summary of the key findings and recommendations for the entire research effort.

5.1 Literature Review and National Survey

The literature review and national survey provided significant insight into alternatives or directions that might be taken to improve the current PENNDOT growth factor estimation program. Both the literature review and national survey indicated that the number of sites in the short-term count program needs to be greatly expanded, while the three-week count duration can be shortened. The new edition of the TMG strongly recommends that short-term HPMS counts be 48-hours in duration. Currently, PENNDOT collects 24-hour counts for the HPMS program. Because of their spatial density, the HPMS counts are a good source of data for growth factor estimation, particularly if they are expanded to 48 hours. The usage of 48-hour counts, and in particular the HPMS counts, is an approach frequently used by other states.

Also, note that the number of growth factors (42) calculated by PENNDOT was not found to be unusually large, and in fact was less than most states. This has two important ramifications for the current program. First, suggestions that the current PENNDOT program is deficient because it simply has too many categories should be dismissed. Second, a larger spatial sample size provides opportunities for expansion of the number of growth factors computed, if such an expansion is ever desired in the future.

5.2 Testing of Spatial versus Temporal Counting Programs

In agreement with the nationwide survey and literature review, the results of this testing indicated that PENNDOT should consider (1) including all short-term HPMS counts in the growth factor estimation process, (2) shortening the duration of the control counts to one week or less, (3) and using more than the last two years in growth rate computations. It was also noted that the control count database contained many errors and should be rigorously inspected and cleaned prior to inclusion in the analysis. Each of these recommendations is discussed in greater detail below.

The results of the testing indicated little if any difference between 21-day count durations and 24-hour count durations for the control counts. It is recommended that the short-term HPMS counts be increased to 48 hours, in compliance with the TMG, and that these also be used for growth factor estimation. The duration of the control counts can be reduced from three weeks to less than one week, depending on human resources, counter battery life, and desired consistency with the HPMS count duration.

A historical database should be used in the estimation of growth factors. In this research, data up to eight years old were used. This provided an advantage over using only the last two (or three) years. The analysis did not address the number of years to use; however, that is likely not a significant item at this point, since most of the data available for growth factor estimation will only be a few years old. As a larger historical database is assembled, this issue should be revisited. Note that in the nationwide survey, of the 36 states responding to the question of "years" considered, the most popular answer was 16 to 20 years, as 16 states utilized this duration.

It is strongly recommended that the count data be diligently inspected and cleaned before use. In this research, significant effort was devoted to cleaning the data before the analysis was performed. Had the data not been cleaned, the results would have been significantly affected. The same could be said for the computation of the growth rates. In addition, it was noted that the control counts rarely had three full weeks in both directions at a particular station. This likely has an effect on the quality of the growth factors that use these counts. It is also another reason to limit the control count duration to a single week, since the current effort rarely produces the desired counts.

Finally, it must be noted that the "HPMS vs. Control Count Comparison" research was limited by the number of ATRs in Allegheny County. There were only four ATRs available for comparison purposes. More ATRs would have provided a more reliable comparison. In addition, there were no rural facilities in Allegheny County. The results are based entirely upon urban facilities. Furthermore, the "Count Duration Comparison" research was limited by the number of ATRs in each of the 42 groups. Most groups had three or less ATRs.

5.3 Cluster Analysis of ATRs

In compliance with the TMG, a cluster analysis of the PENNDOT ATRs was performed. Two key findings were made relative to the existing set of traffic pattern groups.

- The existing set of traffic pattern groups should be retained.
- Four counters (27, 306, 371, and 384) were identified as having some of the attributes of "recreational" facilities. Counter 306 is considered recreational in the existing set of traffic pattern groups. The others may be considered for reclassification as recreational.

A precision analysis was performed for the nine non-recreational traffic pattern groups (excluding the ATRs identified as recreational and ATR 383 which had discrepancies). All TPGs were found to be acceptable.

REFERENCES

Highway Statistics 1998, Report # FHWA-PL-99-017, Federal Highway Administration, Washington, D.C., November 1999.

Traffic Monitoring Guide, 3rd Ed., Report # FHWA PL-95-031, Federal Highway Administration, Washington, D.C., February 1995.

Traffic Monitoring Guide, 4th Ed. (Draft), Report # FHWA PL-01-021, Federal Highway Administration, Washington, D.C., 2001.

APPENDIX A

TECHNICAL MATERIAL SUPPORTING THE LITERATURE REVIEW AND NATIONWIDE SURVEY

Short-Term Count Program Details.

State	Number of Counts	Duration	Frequency
Alabama	7,000	Urban - 48 hours, Rural - one week	3 years
Arkansas	9,000	48 hours	Every year
Colorado	7,000	Urban - 24 hours, Rural - 48 hours	3 years
Delaware	By Special Request Only		
Florida	7,119	no answer	no answer
Georgia	16,000	48 hours	3 year, may start every year
Idaho	350-400	48 hours	2-3 years
Illinois	20,000	no answer	2 years interstate, others 5 years
Iowa	10 - 12,000	24 hours	4 years
Kansas	11,000	48 hours	interstates every year, others every other year
Kentucky	6,876	48 hours	some 3 years, 6 years, and semi-annually
Louisiana	7,800-8,000	48 hours	3 years
Maine	no answer	24 hours	ranging from every other year to every 7 years
Maryland	2,884	48 hours	3 years
Massachusetts	2,424	24 hours	3 years
Michigan	3,540	48 hours	2 year cycle except for ramps in Detroit, which are every 3 yr
Mississippi	9,000	48 hours	3 years
Missouri	10,000	48 hours	3 years

Montana	3,500	36 hours, (some urban 24 hours)	some urban areas every other year, others every 3 years
Nebraska	4,000	24 hours (urban 48 hours)	Every other year
Nevada	4,000	One week	most every year, some every 3 years
New Hampshire	2,000	One week	3 years
New Jersey	3,000	48 hours	3 years
New York	8,000	72 hours	mainly 3 years, some 5 years
North Dakota	hundreds	24 hours	2 years
Oklahoma	1,000	24 hours	2 years (interstates one year, non-interstates one year)
Oregon	6,000	48 hours	3 years
Rhode Island	1,000-1,500	48 hours	3 years
South Carolina	12,000	24 hours	Every year
South Dakota	6,000	24 hours	HPMS every year, others every other year
Tennessee	12,000	24 hours	interstates and SR every year, other ?
Texas	every highway segment, number not determined	48 hours unless manual, in which case it is 24 hours	every year
Utah	3,000	48 hours	3 years
Vermont	600	One week	4 year cycle
Virginia	nearly 100,000	some 24 hr, some 48	some 3 years, some 6-12 years
Washington	3,960 + manual counts	72 hours (4 hour manual counts)	3 years

West Virginia	7,500	48 hours	3 years
Wisconsin	no answer	48 hours	3 years
Wyoming	no answer	24 hours	3 years

APPENDIX B

TECHNICAL MATERIAL SUPPORTING THE TESTING OF SPATIAL VERSUS TEMPORAL COUNTING PROGRAMS

Growth Rates at the ATRs in Allegheny County

ATR	Func Class	TPG	Control Count Group	Interstate or Non?	AADTs (from Annual Report) 1998	1999	2000	Growth Rates 98-99	99-00	98-00
203	14	3	2	Non	20,660	20,719	21,023	0.3%	1.5%	0.9%
375	12	3	2	Non	24,335	22,714	23,990	-6.7%	5.6%	-0.55%
208	11	1	13	Interstate	65,299	64,300	65,332	-1.5%	1.6%	0.05%
309	11	1	13	Interstate	106,981	106,938	106,761	0.0%	-0.2%	-0.1%

Growth Rates at Allegheny County HPMS and Control Count Sites

Control Count Data

Group	Description	98-99 Number of Counters	98-99 Growth Rate	99-00 Number of Counters	99-00 Growth Rate	98-00 Growth Rate
13	Interstate-Urban	2	3.10%	4	3.00%	3.05%
2	Non-Interstate-Urban	16	2.80%	17	-0.50%	1.15%

HPMS Data (Using Unweighted Aggregation), using last two available years only

Group	Description	98-99 Number of Counters	98-99 Growth Rate	99-00 Number of Counters	99-00 Growth Rate	98-00 Growth Rate
13	Interstate-Urban	2	4.07%	4	-0.59%	1.74%
2	Non-Interstate-Urban	56	1.52%	65	2.46%	1.99%

HPMS Data (Using Weighted Aggregation), using last two available years only

Group	Description	98-99 Number of Counters	98-99 Growth Rate	99-00 Number of Counters	99-00 Growth Rate	98-00 Growth Rate
13	Interstate-Urban	2	3.09%	4	-2.88%	0.11%
2	Non-Interstate-Urban	56	2.34%	65	2.79%	2.57%

HPMS Data (Using Unweighted Aggregation), using all years

Group	Description	Number of Counters	Growth Rate
13	Interstate-Urban	4	-0.66%
2	Non-Interstate-Urban	65	1.15%

HPMS Data (Using Weighted Aggregation), using all years

Group	Description	Number of Counters	Growth Rate
13	Interstate-Urban	4	-2.77%
2	Non-Interstate-Urban	65	2.18%

HPMS vs Control Count Data
Absolute Error Calculations

Control Count Data Absolute Error

ATR	ATR Growth	CC Growth	Absolute Error
203	0.9%	1.15%	0.25%
375	-0.55%	1.15%	1.70%
208	0.05%	3.05%	3.00%
309	-0.1%	3.05%	3.15%
		AVERAGE	**2.03%**

HPMS Data Absolute Error
Unweighted Aggregation, All Years Considered

ATR	ATR Growth	HPMS Growth	Absolute Error
203	0.9%	1.15%	0.25%
375	-0.55%	1.15%	1.70%
208	0.05%	-0.66%	0.71%
309	-0.1%	-0.66%	0.56%
		AVERAGE	**0.80%**

HPMS Data Absolute Error
Unweighted Aggregation,
Last Two Available Years Considered

ATR	ATR Growth	HPMS Growth	Absolute Error
203	0.9%	1.99%	1.09%
375	-0.55%	1.99%	2.54%
208	0.05%	1.74%	1.69%
309	-0.1%	1.74%	1.84%
		AVERAGE	**1.79%**

HPMS Data Absolute Error
Weighted Aggregation, All Years Considered

ATR	ATR Growth	HPMS Growth	Absolute Error
203	0.9%	2.18%	1.28%
375	-0.55%	2.18%	2.73%
208	0.05%	-2.77%	2.82%
309	-0.1%	-2.77%	2.67%
		AVERAGE	**2.37%**

HPMS Data Absolute Error
Weighted Aggregation,
Last Two Available Years Considered

ATR	ATR Growth	HPMS Growth	Absolute Error
203	0.9%	2.56%	1.67%
375	-0.55%	2.56%	3.11%
208	0.05%	0.10%	0.05%
309	-0.1%	0.10%	0.20%
		AVERAGE	**1.26%**

Summary - Absolute Errors in Rank Order

0.80% **HPMS, Unweighted Aggregation, All Years Considered**
1.26% **HPMS, Weighted Aggregation, Last Two Available Years Considered**
1.79% **HPMS, Unweighted Aggregation, Last Two Available Years Considered**
2.03% **Control Counts**
2.37% **HPMS, Weighted Aggregation, All Years Considered**

APPENDIX C

TECHNICAL MATERIAL SUPPORTING THE CLUSTER ANALYSIS OF ATRS

Monthly Factor Data

FN CLASS	COUNTER	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	AVG	STD.DEV.	MCOV
1	207	1.3348	1.2778	1.1692	1.0495	0.9525	0.8600	0.7497	0.7465	0.9107	0.9880	1.1390	1.2907	1.0390	0.2051	19.7409
1	216	1.4371	1.2382	1.1433	0.9890	0.9202	0.8771	0.7917	0.8033	0.9728	1.0128	1.0649	1.0955	1.0288	0.1850	17.9792
1	370	1.1656	1.0835	1.0107	0.9937	0.9622	0.9360	0.9532	0.9031	0.9613	0.9819	1.0228	1.0859	1.0050	0.0745	7.4155
1	371	1.4298	1.3331	1.1476	0.9983	0.9542	0.8597	0.7367	0.7749	0.9593	1.0348	1.0174	1.2246	1.0392	0.2120	20.4003
1	372	1.2346	1.2313	1.1106	1.0502	0.9515	0.8872	0.8595	0.8312	0.9212	0.9950	1.0476	1.0786	1.0165	0.1342	13.1992
1	374	1.3984	1.2610	1.0843	1.0130	0.9167	0.8772	0.8256	0.8385	0.9171	0.9247	1.0366	1.2442	1.0281	0.1850	17.9979
1	392	1.3014	1.2238	1.0823	1.0501	0.9288	0.8647	1.1467	0.7924	0.9107	0.9143	0.9310	1.1080	1.0212	0.1551	15.1863
1	393	1.2140	1.2080	1.1014	1.0439	0.9709	0.9022	0.9129	0.8623	0.9558	0.9554	0.9411	1.0774	1.0121	0.1167	11.5339
2	4	1.2975	1.1924	1.1925	1.0112	0.8950	0.8945	0.8716	0.9060	0.9602	0.9045	0.9871	1.0973	1.0175	0.1436	14.1140
2	19	1.1766	1.0988	1.0331	0.9981	0.8963	0.8878	0.9671	0.9539	0.9404	0.9799	1.1265	1.0301	1.0074	0.0898	8.9192
2	24	1.2692	1.1688	1.0746	0.9396	0.8657	0.8948	0.9609	0.9182	0.9511	0.9575	0.9991	1.1616	1.0134	0.1264	12.4693
2	323	1.2623	1.1174	1.0571	0.9921	0.9510	0.9416	0.9605	0.9143	0.9348	0.8939	0.9783	1.1130	1.0097	0.1080	10.6927
2	326	1.1948	1.0810	0.9929	0.9654	0.9032	0.9237	0.9700	0.9460	0.9505	0.9720	1.0440	1.1393	1.0069	0.0898	8.9215
2	334	1.2107	1.0792	1.0076	0.9564	0.9397	0.9186	0.9478	0.9296	0.9375	0.9631	1.0589	1.1412	1.0075	0.0947	9.3973
2	349	1.1085	1.0439	1.0081	0.9990	0.9822	0.9393	0.9938	0.9500	0.9810	0.9792	1.0044	1.0319	1.0018	0.0448	4.4735
2	360	1.1935	1.1254	1.0415	0.9925	0.9382	0.9313	0.9220	0.9188	0.9553	0.9499	1.0225	1.0965	1.0073	0.0904	8.9732
2	363	1.2112	1.1417	1.0622	0.9877	0.9571	0.9217	0.8904	0.8853	0.9431	0.9386	0.9936	1.2091	1.0118	0.1172	11.5796
2	378	1.2093	1.0619	0.9988	0.9647	0.9371	0.9413	0.9242	0.9551	0.9711	0.9763	1.0357	1.0933	1.0057	0.0826	8.2112
6	1	1.2127	1.1513	1.0470	1.0150	0.9643	0.9074	0.8676	0.8986	0.9911	0.9430	1.0038	1.1225	1.0104	0.1067	10.5617
6	2	1.3082	1.1922	1.0464	0.9999	0.9108	0.8824	0.8514	0.8291	0.9424	0.9892	1.0949	1.1951	1.0202	0.1516	14.8590
6	3	1.2724	1.1776	1.0818	0.9645	0.9363	0.9431	0.9321	0.9088	0.9306	0.9040	0.9637	1.1354	1.0125	0.1227	12.1222
6	15	1.1460	1.1039	1.0084	0.9930	0.9539	0.9292	0.9586	0.9409	0.9868	0.9208	1.0138	1.1072	1.0052	0.0753	7.4914
6	27	1.4480	1.3673	1.2079	0.9991	0.9281	0.8933	0.8102	0.8570	0.8888	0.8837	0.9253	1.2392	1.0373	0.2182	21.0376
6	40	1.1096	1.0806	0.9847	0.9693	0.9197	0.9446	0.9838	0.9773	0.9713	0.9726	1.0305	1.0983	1.0035	0.0618	6.1572
6	48	1.2978	1.1853	1.0601	1.0137	0.9049	0.9026	0.8996	0.8746	0.9437	0.9630	1.0527	1.0714	1.0141	0.1283	12.6554
6	51	1.1875	1.1269	1.0361	1.0089	0.9209	0.9341	0.9574	0.9066	0.9729	0.9529	0.9915	1.0868	1.0069	0.0873	8.6670
6	306	1.3219	1.2345	1.2089	1.0862	0.9108	0.8679	0.7214	0.7509	0.9073	1.0107	1.1820	1.3014	1.0420	0.2102	20.1773
6	328	1.1947	1.1228	1.0482	0.9878	0.9186	0.9118	0.9288	0.9492	0.9593	0.9709	1.0142	1.0758	1.0068	0.0881	8.7505
6	367	1.2231	1.1583	1.0601	0.9734	0.9323	0.9281	0.9130	0.9180	0.9441	0.9537	1.0069	1.1000	1.0092	0.1033	10.2394
6	390	1.1644	1.0532	0.9792	0.9739	0.9727	0.9475	0.9592	0.9339	0.9885	0.9850	1.0177	1.0679	1.0036	0.0644	6.4165
6	391	1.3005	1.1469	1.0185	0.9849	0.9586	0.9373	0.9251	0.8820	0.9549	0.9423	0.9942	1.0882	1.0111	0.1165	11.5229
7	5	1.2707	1.1550	1.1334	0.9957	0.8921	0.8946	0.9115	0.8924	0.9338	0.9477	1.0028	1.1386	1.0140	0.1284	12.6583
7	29	1.2586	1.2580	1.1224	1.0415	0.9148	0.9258	0.8870	0.8956	0.9392	0.9565	0.9360	1.0463	1.0151	0.1334	13.1390
7	362	1.1980	1.0935	0.9983	0.9600	0.9393	0.9536	0.9958	0.9609	0.9722	0.9359	1.0098	1.0417	1.0049	0.0759	7.5522
7	364	1.2513	1.1904	1.0799	1.0211	0.9264	0.8747	0.8854	0.8828	0.9470	0.9804	1.0378	1.0809	1.0132	0.1218	12.0257
7	384	1.4399	1.4122	1.2674	1.2238	0.9713	0.8212	0.6223	0.7610	0.9970	1.0169	1.0495	1.1567	1.0616	0.2522	23.7526

7	386	1.1719	1.0937	0.9930	0.9457	0.8941	0.9031	0.9398	0.8838	0.9828	1.0181	1.1006	1.1963	1.0103	0.1075	10.6395
7	388	1.2294	1.1467	1.0157	0.9847	0.9310	0.9147	0.9643	0.9370	0.9626	0.9462	1.0008	1.0590	1.0077	0.0948	9.4127
7	389	1.1801	1.1253	0.9835	0.9527	0.9072	0.9358	0.9816	0.9117	0.9468	0.9679	1.0647	1.1378	1.0079	0.0943	9.3555
8	385	1.2743	1.2000	1.0938	0.9787	0.9215	0.8950	0.9048	0.9020	0.9116	0.9425	1.0192	1.1280	1.0143	0.1300	12.8122
8	387	1.2625	1.1169	1.0231	0.9378	0.9079	0.9007	0.9521	0.8731	0.9606	0.9556	1.0515	1.2168	1.0132	0.1261	12.4416
11	205	1.1766	1.0540	1.0157	0.9974	0.9678	0.9403	0.9615	0.9087	0.9762	0.9639	1.0011	1.0919	1.0046	0.0731	7.2813
11	208	1.1174	1.0486	1.0046	1.0057	0.9741	0.9703	0.9895	0.9278	0.9566	0.9681	0.9981	1.0693	1.0025	0.0528	5.2690
11	210	1.1253	0.9807	0.9756	0.9875	0.9692	0.9486	0.9807	0.9143	1.0950	0.9880	1.0323	1.0414	1.0032	0.0602	6.0045
11	309	1.0515	0.9573	0.9568	1.0132	1.0275	1.0485	0.9850	0.9400	0.9602	1.0320	0.9852	1.0597	1.0014	0.0424	4.2378
11	373	1.2497	1.1157	1.0308	0.9418	0.9551	0.9291	0.9283	0.9050	0.9850	0.9675	1.0072	1.0840	1.0083	0.0995	9.8650
11	376	1.2442	1.1350	1.0704	0.9785	0.9886	0.8897	0.9240	0.8994	0.9696	0.9549	0.9794	1.0769	1.0092	0.1043	10.3370
11	377	1.0967	1.0351	1.0097	1.0246	0.9888	0.9714	1.0002	0.9733	0.9970	0.9904	0.9780	0.9513	1.0014	0.0380	3.7940
11	394	1.2685	1.1340	1.0607	1.0186	0.9514	0.8825	0.9330	0.8515	0.9679	0.9657	1.0154	1.0894	1.0115	0.1151	11.3748
12	206	1.1015	0.9981	0.9525	1.0121	0.9704	0.9895	1.0843	1.0106	0.9455	0.9130	0.9748	1.0830	1.0029	0.0594	5.9271
12	375	1.1305	1.0726	1.0108	1.0255	0.9499	0.9301	0.9717	0.9137	0.9930	0.9757	0.9999	1.0723	1.0038	0.0635	6.3293
14	8	1.0968	1.0412	0.9724	0.9846	0.9489	0.9440	1.0461	1.0138	0.9872	0.9353	1.0010	1.0557	1.0022	0.0500	4.9911
14	203	1.0742	1.0057	0.9686	0.9991	0.9853	0.9398	0.9821	0.9568	1.0203	0.9994	1.0047	1.0824	1.0015	0.0424	4.2318
14	301	1.1259	1.0640	1.0033	0.9971	0.9332	0.8906	0.9548	0.9168	0.9669	0.9848	1.0877	1.1500	1.0063	0.0833	8.2816
14	304	1.2040	1.1240	1.0349	0.9606	0.9433	0.9381	0.9355	0.8931	0.9817	0.9628	1.0185	1.0930	1.0075	0.0917	9.0996
14	330	1.1558	1.0344	0.9774	0.9938	0.9561	0.9342	1.0316	1.0792	0.9969	0.9619	0.9638	0.9544	1.0033	0.0634	6.3214
16	18	1.1772	1.1262	1.0160	0.9617	0.9376	0.9270	0.9615	0.9291	0.9703	0.9731	1.0147	1.0750	1.0058	0.0808	8.0343
16	20	1.1484	1.0784	1.0219	0.9510	0.8998	0.9151	0.9575	0.9642	0.9959	1.0051	1.0559	1.0656	1.0049	0.0730	7.2616
17	379	1.1329	1.0604	0.9945	0.9582	0.9449	0.9628	1.0137	0.8847	0.9710	0.9744	1.0456	1.1143	1.0048	0.0722	7.1866
17	380	1.1716	1.0793	0.9900	0.9498	0.9223	0.9421	1.0076	0.9902	0.9800	0.9540	1.0159	1.0470	1.0042	0.0692	6.8905
17	381	1.0808	1.1095	1.0361	1.0367	0.9293	0.9321	0.9438	0.9536	0.9933	0.9613	1.0198	1.0488	1.0038	0.0605	6.0270
17	382	1.1680	1.1301	1.0205	0.9515	0.9117	0.9118	0.9422	0.9351	0.9800	0.9743	1.0608	1.0983	1.0070	0.0878	8.7236
19	383	1.1693	1.1539	1.0342	0.9853	0.8989	0.9137	0.9606	0.9316	0.9540	0.9635	1.0437	1.0799	1.0074	0.0898	8.9157

SAS Program to Perform Cluster Analysis Using Only Monthly Factors for Clustering

```
data penndot;
input fc counter m1-m12 aaa sss cv;
If counter = 27 then delete;
If counter = 306 then delete;
If counter = 371 then delete;
If counter = 383 then delete;
If counter = 384 then delete;
cards;
INSERT DATA HERE
run;
proc cluster
 method = Ward;
 var m1-m12;
 ID counter;
Run;
```

Results of the Cluster Analysis Using Only Monthly Factors

The CLUSTER Procedure
Ward's Minimum Variance Cluster Analysis

Eigenvalues of the Covariance Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	0.01888400	0.01519363	0.6158	0.6158
2	0.00369037	0.00133330	0.1203	0.7361
3	0.00235707	0.00064649	0.0769	0.8130
4	0.00171058	0.00053996	0.0558	0.8687
5	0.00117062	0.00006415	0.0382	0.9069
6	0.00110646	0.00057584	0.0361	0.9430
7	0.00053062	0.00012003	0.0173	0.9603
8	0.00041059	0.00005315	0.0134	0.9737
9	0.00035744	0.00015017	0.0117	0.9853
10	0.00020726	0.00001947	0.0068	0.9921
11	0.00018779	0.00013315	0.0061	0.9982
12	0.00005464		0.0018	1.0000

Root-Mean-Square Total-Sample Standard Deviation = 0.050553
Root-Mean-Square Distance Between Observations = 0.247659

Cluster History

NCL	------Clusters Joined-------		FREQ	SPRSQ	RSQ
57	370	205	2	0.0006	.999
56	360	328	2	0.0008	.999
55	362	380	2	0.0008	.998
54	326	334	2	0.0009	.997
53	373	304	2	0.0009	.996
52	349	208	2	0.0011	.995
51	18	382	2	0.0012	.994
50	CL56	51	3	0.0013	.992
49	48	364	2	0.0014	.991
48	CL57	390	3	0.0015	.989
47	5	385	2	0.0017	.988
46	CL50	367	4	0.0018	.986
45	CL54	389	3	0.0018	.984
44	CL48	375	4	0.0018	.982
43	323	3	2	0.0020	.980
42	CL46	388	5	0.0021	.978
41	376	394	2	0.0022	.976
40	40	20	2	0.0025	.974
39	15	CL51	3	0.0029	.971
38	391	CL41	3	0.0030	.968
37	206	8	2	0.0031	.965
36	CL45	378	4	0.0033	.961
35	393	29	2	0.0033	.958
34	363	1	2	0.0034	.954
33	CL42	CL53	7	0.0035	.951
32	386	301	2	0.0036	.947
31	CL52	203	3	0.0037	.944
30	24	CL47	3	0.0041	.940
29	CL44	379	5	0.0043	.935
28	372	CL49	3	0.0046	.931
27	CL39	CL40	5	0.0046	.926
26	377	330	2	0.0061	.920
25	CL27	381	6	0.0062	.914
24	CL43	CL38	5	0.0072	.907
23	374	2	2	0.0074	.899
22	CL25	CL55	8	0.0076	.892
21	CL29	CL31	8	0.0077	.884
20	CL36	CL32	6	0.0081	.876
19	4	CL30	4	0.0082	.868
18	CL34	387	3	0.0087	.859

Cluster History

NCL	------Clusters Joined-------		FREQ	SPRSQ	RSQ
17	19	CL22	9	0.0094	.850
16	CL21	210	9	0.0105	.839
15	CL28	CL35	5	0.0113	.828
14	CL24	CL33	12	0.0115	.816
13	CL26	CL37	4	0.0129	.803
12	216	CL23	3	0.0141	.789
11	CL14	CL18	15	0.0142	.775
10	CL15	CL19	9	0.0174	.758
9	309	CL13	5	0.0182	.739
8	207	CL12	4	0.0194	.720
7	CL17	CL20	15	0.0201	.700
6	CL16	CL7	24	0.0318	.668
5	CL10	CL11	24	0.0405	.628
4	CL5	392	25	0.0463	.581
3	CL6	CL9	29	0.0660	.515
2	CL8	CL4	29	0.1297	.386
1	CL2	CL3	58	0.3856	.000

SAS Program to Perform Cluster Analysis Using Functional Class and Monthly Factors for Clustering

```
data penndot;
input fc counter m1-m12 aaa sss cv;
If counter = 27 then delete;
If counter = 306 then delete;
If counter = 371 then delete;
If counter = 383 then delete;
If counter = 384 then delete;
if fc=1 then fcode = 10
if fc=2 then fcode = 20
if fc=6 then fcode = 30
if fc=7 then fcode = 40
if fc=8 then fcode = 50
if fc=9 then fcode = 60
if fc=11 then fcode = 110
if fc=12 then fcode = 120
if fc=14then fcode = 130
if fc=16 then fcode = 140
if fc=17 then fcode = 150
if fc=19 then fcode = 160
cards;
INSERT DATA HERE
run;
proc cluster
  method = Ward;
  var fcode m1-m12;
  ID counter;
Run;
```

Results of the Cluster Analysis Using Functional Class and Monthly Factors for Clustering

The CLUSTER Procedure
Ward's Minimum Variance Cluster Analysis

Eigenvalues of the Covariance Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	2507.23576	2507.22323	1.0000	1.0000
2	0.01253	0.00885	0.0000	1.0000
3	0.00367	0.00136	0.0000	1.0000
4	0.00231	0.00060	0.0000	1.0000
5	0.00171	0.00054	0.0000	1.0000
6	0.00117	0.00007	0.0000	1.0000
7	0.00110	0.00060	0.0000	1.0000
8	0.00050	0.00009	0.0000	1.0000
9	0.00041	0.00005	0.0000	1.0000
10	0.00036	0.00015	0.0000	1.0000
11	0.00021	0.00002	0.0000	1.0000
12	0.00018	0.00013	0.0000	1.0000
13	0.00005		0.0000	1.0000

Root-Mean-Square Total-Sample Standard Deviation = 13.88763
Root-Mean-Square Distance Between Observations = 70.81327

Cluster History

NCL	------Clusters Joined-------		FREQ	SPRSQ	RSQ
57	326	334	2	0.0000	1.00
56	51	328	2	0.0000	1.00
55	205	208	2	0.0000	1.00
54	CL56	367	3	0.0000	1.00
53	373	376	2	0.0000	1.00
52	CL57	378	3	0.0000	1.00
51	362	388	2	0.0000	1.00
50	18	20	2	0.0000	1.00
49	15	40	2	0.0000	1.00
48	CL49	390	3	0.0000	1.00
47	CL53	394	3	0.0000	1.00
46	323	360	2	0.0000	1.00
45	3	391	2	0.0000	1.00
44	5	364	2	0.0000	1.00
43	386	389	2	0.0000	1.00
42	1	CL54	4	0.0000	1.00
41	8	203	2	0.0000	1.00
40	380	382	2	0.0000	1.00
39	372	393	2	0.0000	1.00
38	379	CL40	3	0.0000	1.00
37	CL45	48	3	0.0000	1.00
36	CL46	363	3	0.0000	1.00
35	301	304	2	0.0000	1.00
34	24	CL36	4	0.0000	1.00
33	CL38	381	4	0.0000	1.00
32	385	387	2	0.0000	1.00
31	CL44	29	3	0.0000	1.00
30	210	377	2	0.0000	1.00
29	19	CL52	4	0.0000	1.00
28	CL42	CL37	7	0.0000	1.00
27	CL55	CL30	4	0.0000	1.00
26	CL41	330	3	0.0000	1.00
25	CL29	349	5	0.0000	1.00
24	216	374	2	0.0000	1.00
23	206	375	2	0.0000	1.00
22	CL27	309	5	0.0000	1.00
21	4	CL34	5	0.0000	1.00
20	CL51	CL43	4	0.0000	1.00
19	370	CL39	3	0.0000	1.00

SAS Programs to Perform Cluster Analyses of Functional Classes 6, 7, and 8

Cluster analysis for Rural Minor Arterials (North and Central)

```
data penndot;
input fc counter m1-m12 aaa sss cv;
If counter = 27 then delete;
If counter = 306 then delete;
If counter = 371 then delete;
If counter = 383 then delete;
If counter = 384 then delete;
if fc<6 then delete;
if fc>6 then delete;
cards;
INSERT DATA HERE
run;
proc cluster
 method = Ward;
 var m1-m12;
 ID counter;
Run;
```

Cluster analysis for Rural Collectors (7,8) (North and Central)

```
data penndot;
input fc counter m1-m12 aaa sss cv;
If counter = 27 then delete;
If counter = 306 then delete;
If counter = 371 then delete;
If counter = 383 then delete;
If counter = 384 then delete;
if fc<7 then delete;
if fc>8 then delete;
cards;
INSERT DATA HERE
run;
proc cluster
 method = Ward;
 var m1-m12;
 ID counter;
Run;
```

Cluster analysis for <u>Rural Minor Arterials and Collectors (6,7,8) (North and Central)</u>

```
data penndot;
input fc counter m1-m12 aaa sss cv;
If counter = 27 then delete;
If counter = 306 then delete;
If counter = 371 then delete;
If counter = 383 then delete;
If counter = 384 then delete;
if fc<6 then delete;
if fc>8 then delete;
cards;
INSERT DATA HERE
run;
proc cluster
 method = Ward;
 var m1-m12;
 ID counter;
Run;
```

SAS Results from Cluster Analyses of Functional Classes 6, 7, and 8
Rural Minor Arterials
The CLUSTER Procedure
Ward's Minimum Variance Cluster Analysis

Eigenvalues of the Covariance Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	0.01009200	0.00782612	0.6403	0.6403
2	0.00226588	0.00110193	0.1438	0.7841
3	0.00116395	0.00019677	0.0739	0.8580
4	0.00096719	0.00023743	0.0614	0.9193
5	0.00072976	0.00041666	0.0463	0.9656
6	0.00031309	0.00016889	0.0199	0.9855
7	0.00014420	0.00010347	0.0091	0.9946
8	0.00004073	0.00000095	0.0026	0.9972
9	0.00003978	0.00003577	0.0025	0.9997
10	0.00000401	0.00000401	0.0003	1.0000
11	0.00000000	0.00000000	0.0000	1.0000
12	-.00000000		-0.0000	1.0000

Root-Mean-Square Total-Sample Standard Deviation = 0.036241
Root-Mean-Square Distance Between Observations = 0.177542

Cluster History

NCL	------Clusters	Joined-------		FREQ	SPRSQ	RSQ
10		51	328	2	0.0157	.984
9	CL10		367	3	0.0220	.962
8		15	40	2	0.0306	.932
7	CL8		390	3	0.0346	.897
6		3	391	2	0.0402	.857
5		1	CL9	4	0.0495	.807
4	CL6		48	3	0.0653	.742
3	CL5		CL4	7	0.1111	.631
2	CL3		2	8	0.2131	.418
1	CL2		CL7	11	0.4179	.000

Rural Collectors

The CLUSTER Procedure

Ward's Minimum Variance Cluster Analysis

Eigenvalues of the Covariance Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	0.01154144	0.00605300	0.5810	0.5810
2	0.00548844	0.00405861	0.2763	0.8573
3	0.00142983	0.00088732	0.0720	0.9293
4	0.00054251	0.00006770	0.0273	0.9566
5	0.00047481	0.00018433	0.0239	0.9805
6	0.00029047	0.00022948	0.0146	0.9951
7	0.00006099	0.00002486	0.0031	0.9982
8	0.00003613	0.00003613	0.0018	1.0000
9	0.00000000	0.00000000	0.0000	1.0000
10	0.00000000	0.00000000	0.0000	1.0000
11	-.00000000	0.00000000	-0.0000	1.0000
12	-.00000000		-0.0000	1.0000

Root-Mean-Square Total-Sample Standard Deviation = 0.040686
Root-Mean-Square Distance Between Observations = 0.199322

Cluster History

NCL	------Clusters Joined-------		FREQ	SPRSQ	RSQ
8	5	385	2	0.0182	.982
7	362	388	2	0.0266	.955
6	CL8	364	3	0.0408	.914
5	386	389	2	0.0425	.872
4	CL5	387	3	0.0622	.810
3	CL6	29	4	0.1023	.707
2	CL7	CL4	5	0.2283	.479
1	CL3	CL2	9	0.4791	.000

Rural Minor Arterials and Collectors

The CLUSTER Procedure

Ward's Minimum Variance Cluster Analysis

Eigenvalues of the Covariance Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	0.00862056	0.00416116	0.5021	0.5021
2	0.00445940	0.00311892	0.2597	0.7618
3	0.00134048	0.00031759	0.0781	0.8399
4	0.00102289	0.00025930	0.0596	0.8995
5	0.00076359	0.00041194	0.0445	0.9440
6	0.00035165	0.00010249	0.0205	0.9645
7	0.00024916	0.00006265	0.0145	0.9790
8	0.00018652	0.00010488	0.0109	0.9898
9	0.00008164	0.00002506	0.0048	0.9946
10	0.00005658	0.00002454	0.0033	0.9979
11	0.00003204	0.00002778	0.0019	0.9998
12	0.00000426		0.0002	1.0000

Root-Mean-Square Total-Sample Standard Deviation = 0.037825

Root-Mean-Square Distance Between Observations = 0.185304

Cluster History

NCL	------Clusters Joined-------		FREQ	SPRSQ	RSQ
19	48	364	2	0.0073	.993
18	51	328	2	0.0076	.985
17	CL18	388	3	0.0088	.976
16	5	385	2	0.0089	.967
15	CL17	367	4	0.0116	.956
14	15	40	2	0.0148	.94
13	390	362	2	0.0153	.926
12	3	CL16	3	0.0193	.907
11	386	389	2	0.0207	.886
10	CL14	CL13	4	0.0207	.865
9	1	391	2	0.0253	.840
8	CL11	387	3	0.0303	.809
7	CL9	CL15	6	0.0337	.776
6	CL12	CL19	5	0.0443	.731
5	CL6	29	6	0.0532	.678
4	2	CL5	7	0.0800	.598
3	CL7	CL10	10	0.1157	.483
2	CL3	CL8	13	0.1442	.338
1	CL2	CL4	20	0.3384	.000